Nonparametric Estimation of Distributions in Large-p, Small-n Settings

Jeffrey D. Hart

Texas A&M University

E-Mail: hart@stat.tamu.edu

Abstract: A distinguishing feature of a number of modern problems, including microarray analyses, is that the observed data consist of a large number, p, of quite small data sets. Nonparametric estimation of an error distribution common to all p data sets is very challenging in such settings, being akin to deconvolution. A canonical example of the models of interest is a random effects model in which means of data sets vary randomly, but otherwise all data are identically distributed. There are two distributions to estimate in this model: the common within-data-sets distribution, and the distribution of means. Both old and recent results on nonparametric estimators in this canonical model will be reviewed. We then consider a location-scale model in which, conditional on a data set, the data have a distribution of the form $F((x-\mu)/\sigma)$, where μ and σ are the mean and standard deviation for the data set, and the joint distribution G of (μ, σ) across all data sets is arbitrary. We propose the use of Wolfowitz minimum distance methodology to obtain nonparametric estimates of F and G. A random search algorithm is used to approximate such estimates in the location-scale model. Simulation studies and a microarray example help to illustrate the efficacy of our approach.